Graphics, Vision and Image Processing Journal, ISSN 1687-398X, Volume 17, Issue 2, ICGST LLC, Delaware, USA, Nov. 2017

www.icgst.com

Video Compression using Efficient Encoding Techniques for Low Bit Rate Applications

Poorva Waingankar[1], S.M. Joshi[2]
1.Assoc.Professor, Thakur College of Engg. &Technology, Mumbai, India.
2.Professor, Vidyalankar Institute of Technology, Mumbai, India.
pwaingankar@gmail.com,
http://www.tcetmumbai,in

Abstract
This paper presents use of Accordion technique along with modified Run Length Encoding for video compression, which consists of exploiting the high amount of temporal redundancies present in videos by converting them to spatial redundancy and using 2D DCT. The Video compression steps are either optimized or completely revamped to meet the compression and video quality requirement in mobile application. This technique is less complex to suit lower end CPUs and achieves a very good compression ratio to suit the narrow bandwidth environments of wireless networks, without compromising on the quality of the video.

Keywords:*H.264, Accordion, Run Length Encoding(RLE), Discrete Cosine Transform(DCT), Huffman Encoding*

Nomenclature

DCT	Discrete Cosine Transform
Mbps	Mega Bits per second
MSE	Mean Square Error
PSNR	Peak Signal to Noise Ratio
QP	Quantization Parameter
RLE	Run Length Encoding

1. Introduction
The technological development in multimedia industry over the past decade has enabled widespread usage of internet based applications and smart phones. Out of the various types of media, video transmission and reception through wireless networks is important in the context of the universal access. The increase in communication speed, computing power and availability of computer storage facilities, has led to a new age of multimedia applications. Various applications such as mobile messaging, video conferencing, use of social networking sites etc. require use of multimedia on large scale Although Wireless communications technologies have been evolving rapidly, the available bandwidth is still of great value and so video coding at ultralow bitrates plays an important role in the development of convergent and interoperable video based multimedia services. These applications need storage of high-quality data, reliable transmission and ease of access to content. The volume of data generated by digitizing a video signal is very large for most transmission systems. Therefore, digital video compression is an important aspect in the realization of these applications. The demand for quality, performance and limitations of available transmission capabilities is necessary to be fulfilled by digital video compression techniques. An efficient and well designed video compression system gives significant performance advantages for visual communication at both low and high transmission bandwidths.

The process of transmission and reception of digital video from source to its destination involves many stages. The most important process is compression (encoding) and decompression (decoding). In this the bandwidth-intensive ‘raw’ digital video is reduced to a manageable size for transmission or storage, then reconstructed for display. The proper compression and decompression process can provide better image quality, greater reliability and/or more flexibility. Therefore, researchers have keen interest in the continuing development and improvement of video compression and decompression methods involving various innovative techniques.

In a typical video, often the temporal redundancies are found to be more relevant than spatial one. In the current video compression techniques, these redundancies are not fully exploited. It is possible to achieve more efficient compression by exploiting these redundancies in the temporal domain. In most of the techniques motion estimation and compensation techniques are usually employed to exploit temporal redundancies. It is observed that the motion estimation process is computationally intensive and its real time implementation is difficult as well. Considering current trends and developments in multimedia applications over internet and mobile communication, an effective algorithm which can fully exploit the redundancy would help to reduce the overall bit rate of transmission/reception.

2. Video Compression Fundamentals

An uncompressed video produces an enormous amount of data and need more than 100s of Mbps bandwidth. Such amount of data causes extremely high computational demands even with powerful computing systems. Hence data compression is an important aspect for managing such data. There are mainly two categories of compression; lossy and lossless. In lossy methods; Transform based coding, Vector quantization, block truncation etc. are used whereas, Run Length Encoding, Huffman Coding, Predictive Coding are used for lossless compression.

The lossless compression retains the original data retaining individual image sequences remain the same, hence compression rate is smaller in this case. The "lossy" compression methods remove image and sound information that is unlikely to be noticed by the viewer, thereby volume of data is significantly decreased. There is always a trade-off between data size and the quality. The higher the compression ratio, lower the size and the quality too. The encoding and decoding process also needs computational resources which need to be taken into consideration. The digital video contains a great deal of redundancy which is categorized in three types as given below:

- Spatial redundancy, which is due to the correlation or dependence between neighbouring pixel values
- Spectral redundancy, which is due to the correlation between different colour planes or spectral bands
- Temporal redundancy, which is present because of correlation between different frames in videos.

The spatial redundancy is reduced by registering differences between parts of a single frame; this is known as intraframe compression and is closely related to image compression. Likewise, temporal redundancy can be reduced by registering differences between frames; this is known as interframe compression, including motion compensation and other techniques. Hence for effective video compression, both interframe and intraframe techniques are used. The typical Video Compression system is shown in Figure.1.

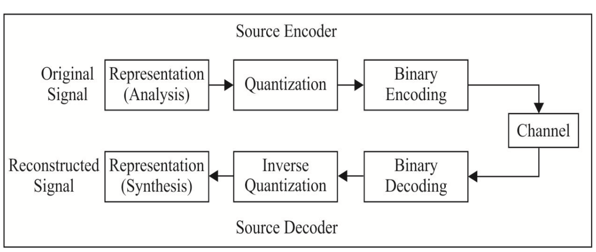

Figure1.Typical Video Compression Scheme

3. Brief Literature Review

The digital video compression technologies have become an integral part of visual information transmitted/received through wired and wireless networks over last one and a half decades. Various standards have been developed for this purpose which define a specific bit stream syntax, imposes very limited constraints on the values of that syntax, and define a limited-scope decoding process. Video codecs are primarily characterized in terms throughput of the channel, distortion of the decoded video, delay and complexity (in terms of computation, memory capacity, and memory access requirements). The intent is for every decoder that conforms to the standard to produce similar output when given a bit stream that conforms to the specified constraints. Thus, these video coding standards are written primarily only to ensure interoperability (and syntax capability), not to ensure quality. This limitation of scope permits maximal freedom to optimize the design of each specific product (balancing compression quality, implementation cost, time to market, etc.). It provides no guarantees of end-to-end reproduction quality, as it allows even crude encoding methods to be considered in conformance with the standard [1][2].

To obtain highly compressed videos without compromising visual quality and to make cost performance trade-offs best suited to applications, researchers have proposed different methods. The multi-objective optimization technique used as a mean for multi-criteria decision making [3]. In which quantization Parameter (QP) controls the tradeoff between quality and bit rate in the sense that a QP increment by 1 results in 12.5% reduction of bit-rate. For network related constraints, optimization algorithm referred to as the Network State Dependent Video Compression Rate (NSDVCR), which determines the compression rates depending on the video characteristics and the network condition is proposed [4].

The possibility of dynamic frame skipping to achieve even higher video compression for low bit rate applications less than 16KbpS is explored by researchers [5]. Motion compensation is very important step in video compression ,so by using control grid interpolation for block based motion compensation, like other interframe compression techniques, produces an approximation of a frame by reusing data contained in the frame's predecessor[6] and in another technique i.e. overlapped block motion compensation is proposed that [7], for each block in the current frame a matching block is found in the past frame and if suitable, its motion vector is substituted for the block during transmission. Depending on the search threshold some blocks will be transmitted in their entirety rather than substituted by motion vectors. The problem of finding the most suitable block in the past frame is known as the block matching problem.

Videos with less motion elements contain high level of temporal redundancy. To avoid the complex computational step of motion estimation and compensation, a new low complexity DCT based video compression method is proposed where Accordion representation converts 3D video content by a 2D image, which allows exploiting the redundancy for high compression [8].

In the subsequent section, use of Accordion representation along with improved Huffman dictionary and modified RLE for Video is presented.

incorporated into the model, it can be shown that significant improvements in the performance of the algorithm can be realised. Moreover, the simplicity and

the efficiency of dynamic pose tracking techniques succeeded to improve the robot pose estimation process.

4. Proposed Methodology

The video signal has high temporal redundancies due to the high correlation between successive frames. It is possible to achieve more efficient compression by exploiting more and more the redundancies in the temporal domain. The proposed method consists of projecting temporal redundancy of each group of pictures into spatial domain to be combined with spatial redundancy in one representation with high spatial correlation i.e. by using **Accordion representation**. The Accordion representation provides a symmetric encoder-decoder design, avoiding the motion compensation step and reduces blocking artifacts. The Accordion representation of any video acts like a preprocessing technique for DCT to achieve a very good amount of energy compaction. The flow chart of an implementation of proposed algorithm is shown in Figure.2.

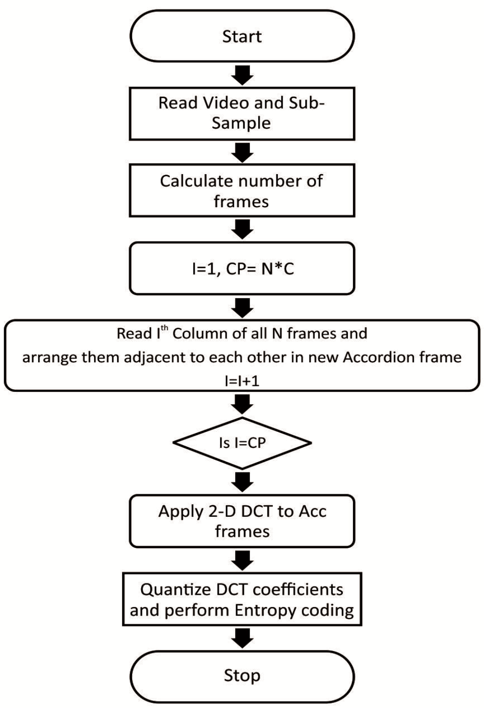

Figure 2: Flow Chart of algorithm

Initially for a video frame, Sub-sampling is implemented by calculating the average pixel value for each group of several pixels, and then substituting this average in the appropriate place in the approximated image. In general, whenever sub sampling is done at the encoder, the decoder has to reconstruct the original picture with some approximation by using a technique called pixel doubling. But in mobile based applications, since the screen resolution is less due to small size, this pixel doubling step is avoided which reduces the decoder complexity.

After considering various factors like the compression percentage, the computational complexity and picture clarity, bilinear interpolation method is found to give the best performance in picture clarity with moderate complexity. After being read into matrix, the input video is divided into several groups with each group consisting of N number of frames where N is the number of frames played in the video per second i.e. fps of the input video. This group having similar temporal frames are gathered into one stretched frame (2 dimensions) by reading each column of every frame subsequently.

The final step consists of coding the obtained frame. The image obtained from the previous steps is now divided into blocks of size 8 x 8, which are then transformed using an 8 x 8 forward DCT. The top-left coefficient in the 2-D DCT array is referred to as the DC coefficient and is proportional to the average brightness of the spatial block. The low-frequency coefficients in the top-left comer of the array have larger values than the higher-frequency coefficients. The transform coefficients are then quantized as per their statistical properties. Most of the energy is concentrated in the low frequency coefficients and hence the higher frequency coefficients which are the least important are harsh quantized or forcibly reduced to zero to avoid any further processing. The Quantization table is designed to provide the most visually correct reconstruction Image. It is designed according to the perceptual importance of the DCT coefficients under the intended viewing conditions. The quality and bit rate of an encoded image can be varied by changing this array. The quantization of AC coefficients creates many zeros, especially at higher frequencies which can be coded efficiently.

The following relation is used for quantization.

$$QDCT = round\ [(8*DCT)/scale*Q) \tag{1}$$

Where, *DCT* is the DCT coefficients, Scale is the scaling factor, Q is the corresponding element of the quantization matrix.

The 2-D array of the DCT coefficients is now formatted into a 1-D vector using a zigzag reordering. Hence the 8 x 8 DCT matrix is now converted to a one dimensional array of 64 coefficients. These 64 numbers are collected by scanning the matrix in zigzag fashion. This rearranges the coefficients in approximately decreasing order of their average energy (as well as in order of increasing spatial frequency) with the aim of creating large runs of zero values since it produces a string of 64 numbers that starts with some non-zeros and typically ends with many consecutive zeros. These runs of zeros are further compressed efficiently using the modified run length encoding procedure.

When the two DC coefficients belonging consecutive DCT matrices have a large difference every such unique difference leads to one unique symbol in the Huffman dictionary in turn leading to many code words which defeats the purpose of compression. To resolve this issue, difference between thee coefficients is coded digit wise with ten unique symbols, thereby code words consequently leading to a much smaller Huffman dictionary. This approach has tremendously reduced the dictionary size and increased the compression ratio.

While carrying out the compression for different videos, it is observed that apart from the number zero, there are very few symbols which have frequent repetitions and hence conventional RLE is not suitable here. This

problem is resolved in the following manner. After analysing the input stream of quantized DCT coefficients in the modified RLE,

a. There is no Run Length Encoding for non-zero elements
b. RLE for all zeros encountered until the last non-zero element
c. Once the last non-zero element is encountered, all the remaining zeros are replaced by special end–of-block (EOB) code with 2 'Zeros'.
Upon reception of the EOB signal, the receiver automatically sets all the remaining coefficients along the zigzag scan to zero. For decoding bit stream, exactly reverse process is carried out step by step. Once the Accordion frame is reconstructed, the MSE and PSNR which are the metrics for reconstructed video quality were calculated using the following relations.

$$PSNR = 10 \log_{10} (Max^2/MSE) \tag{2}$$

Where *Max* is the maximum possible intensity in the image (e.g. 255 for a sample precision of 8 bits), and the Mean Square Error (MSE) is given by:

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j)-K(i,j)\right]^2 \tag{3}$$

Where the number of rows and columns in the image are *m* and *n* respectively.
I(i,j) is the intensity of a pixel at position (i,j) in the original Accordion image, while *K(i,j)* is the value of the corresponding pixel in the compressed and reconstructed Accordion image. The compression in percent is given by;

$$\%C = \frac{(\text{Size of Ori. Video} - \text{Size of Compressed Video})}{\text{Size of Ori. Video}} \tag{4}$$

5.Results

After applying Accordion principle to frames of input video, a stretched frame is formed as shown in Figure 3. This is constructed from 4 sample frames. It can be observed that the temporal redundancies present among the four sample frames is converted to spatial redundancies in the resulting Accordion frame. This step acts as the preprocessing tool to make the 2D DCT very efficient.

Figure 3. Stretched Accordion Frame

After applying 2D-DCT and quantization, it is observed. that the dictionary size reduces to a great extent by using modified RLE and efficient handling of DC coefficients. which is shown in Figure 4. Table 2 shows that for the case of 10 frames, the average length of code words reduces from 3.7375 in conventional technique to 2.877 in improvised technique.

Table 2. Codeword-length with improved RLE and DC

Average Length of Code words					
Number of Frames	2	4	6	8	10
RLE & DC	3.6556	3.6353	3.6893	3.7314	**3.7375**
RLE &Improved DC	3.1162	3.1919	3.1603	3.2349	3.2021
Modified RLE & DC	2.9048	2.8795	2.853	2.8981	**2.8777**

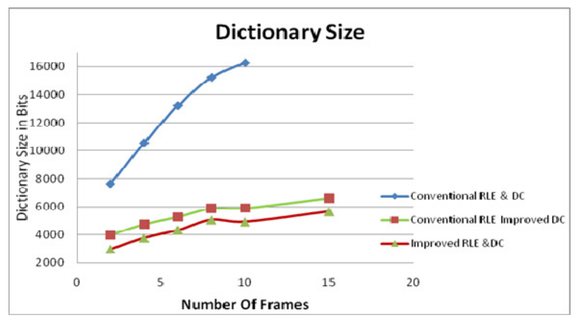

Figure 4. Dictionary Size for RLE and DC Coefficients

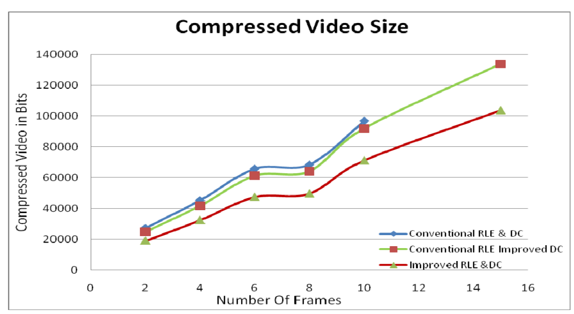

Figure 5. Comparison of compressed Video Size

Finally, the reduction in the size of the compressed video by using the proposed algorithm can be observed from Figure 5. The comparison based on PSNR is shown in Figure 6. It is very much evident that in spite of using different techniques to increase the compression, there is no or very little change in the PSNR of the reconstructed video and it is maintained at around 48 dBs. This indicates that the reconstructed video is of very good quality.

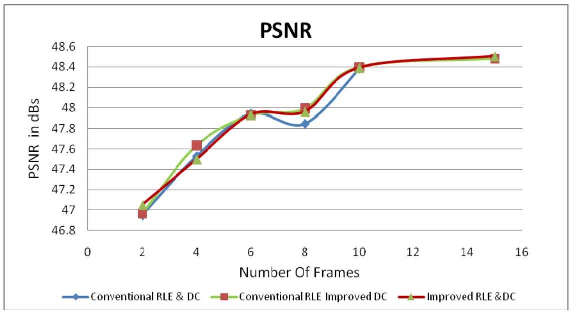

Figure 6. PSNR comparison

Further,the scale of quantization was increased from 1 to 5 for 15 frames of input video.The result of varying the quantization scale is depicted in the table 3(a)and(b).It is observed that by increasing the scale of quantization, the bit stream and the dictionary size of the compressed video reduces considerably while maintaining a good PSNR.

Table 3.(a) PSNR vs Quantization Scale

.

PSNR	
Quantization Scale=1	48.5058
Quantization Scale=2	47.0732
Quantization Scale=3	46.0427
Quantization Scale=4	44.0471
Quantization Scale=5	42.1601

Table 3(b) Bitstream vs Quantization Scale

Bit Stream Size	
Quantization Scale=1	98339
Quantization Scale=2	89466
Quantization Scale=3	83506
Quantization Scale=4	77998
Quantization Scale=5	72108

Since the PSNR is in the acceptable range for even the quantization scale of 5, depending on the required picture quality one can chose the scale and the compression ratio. In the next section, conclusive remarks are given.

6. Conclusions

In this paper, use of Accordion technique for video compression is presented. This technique consists of exploiting the high amount of temporal redundancies present in videos by converting them to spatial redundancy and using 2D DCT. Also, the conventional approaches related to Zigzag processing and Run Length Encoding are re-designed to get a further optimized Huffman dictionary. All the algorithms are developed in MATLAB environment. On comparing the conventional techniques and the proposed algorithm, a significant reduction of 60% in size of Huffman dictionary and 25% reduction in code word length are found by processing the DC components in this unique way. This results in a significant reduction in the size of compressed video while maintaining the PSNR at the same level (around 48db). The subjective quality of video is observed by varying the quantization scale. The quantization scale is varied from 1 to 5, and it has been observed that even with a scale of 5 the reconstructed video is of good visual quality. This technique can be effectively used for slow moving objects such as video conferencing, surveillance etc. However, the rest of optimization techniques will yield a significant additional compression without losing the video quality measured in PSNR.

References

[1] Thomas Wiegand and Gary J. Sullivan," —The H.264/AVC Video Coding Standards"; *IEEE Signal Processing Magazine*, March 2007.

[2] Gary j. Sullivan, Senior member, IEEE and Thomas Wiegand, "Video Compression—From Concepts to the H.264/AVC Standard", 0018-9219 © 2005 IEEE.

[3] Ashraf A. Al-Najdawi, Roy S. Kalawsky," Multiple objective optimization framework for Video compression and transmission,"978-1-4244-18763/08©2008 IEEE.

[4] Xiaoling Qiu, Haiping Liu, Dipak Ghosal and Biswanath Mukherjee, John Benko, Wei Li and Rashmi Bajaj, "Adaptive Video Compression Rate Optimization in Wireless Access Networks",978-1-4244-4487-8/09 ©2009 IEEE.

[5] Antonio Silva and Antonio Navarro," Ultra Low Bitrate Video Coding ",0-7803-8920-4/05©2005 IEEE.

[6] G.J. Sullivan and R.L. Baker,"Motion compensation for video compression using control grid interpolation", Proc. IEEE Int. Conf. Acoustics Speech, and Signal Processing (ICASSP), 1991, pp 2713–2716.

[7] G.J.Sullivan,"Multi-hypothesismotion compensation for low bit rate video coding", Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 1993, pp 437- 440.

[8] Tarek Ouni, Walid Ayedi, Mohamed Abid, "New low complexity DCT based video compression Method",978-1-4244-2937-0/09 © 2009 IEEE.

Biographies

Ms. Poorva Waingankar is currently working as Associate Professor in Electronics Engineering Department at Thakur College of Engg. &Technology, Mumbai, India and pursuing Ph.D. in Electronics Faculty at Swami Ramanand Tirth Marathwada University, Nanded, India.

Dr.Sangeeta Joshi is currently Professor and Technical Advisor at Vidyalankar Institute of Technology, Mumbai, India. She has published number of research papers in reputed journals and received many awards. Her major fields of interest are, Embedded Systems and PC Based Instrumentation, VLSI Design, Reconfigurable Computing, Sensor Network, Simulation and Modeling of Nanodevices.